Actionable Ethics for Data Scientists

Katie Wetstone (Data Scientist, DrivenData)
Isha Shah (Data Scientist, DrivenData)

Agenda

1. Introduction to data ethics
    a. Why data scientists should learn about ethics
    b. Overview of Deon, an ethics checklist for data scientists

   Break

2. Group activity: Qualitative case studies

   Break

3. Hands-on coding exercise: Eviction data case study

Goal: Provide a checklist-based framework for integrating ethics considerations into data science workflows

Who are we?

DRIVENDATA

Data Science + Social Impact

Data Science Competitions • Direct Client Engagements • Open Source Projects

drivendata.org

https://github.com/drivendataorg

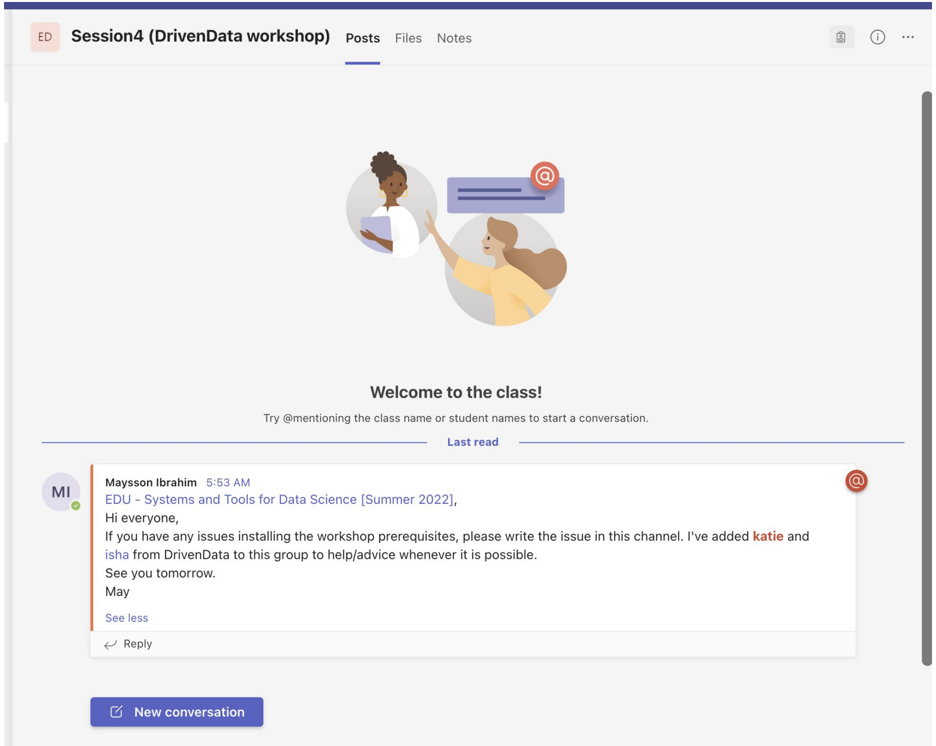
[@drivendataorg](https://twitter.com/drivendataorg)

Who are you?

What is your background?

Post in the Teams channel with either:

- What motivated you to get a masters in data science, or
- What domain your professional background is in



Introduction to data ethics

Why do data scientists need to learn about ethics?

ethics */ˈɛθɪks/ (plural noun)*

Moral principles that govern a person's behaviour or the conducting of an activity.

Data science, machine learning, and AI are playing a larger role in industries that have a huge effect on people's daily lives.

Machine Learning Breakthroughs Have Sparked the AI Revolution

With innumerable data and hyperefficient artificial intelligence, this takeover is coming at lightning speed

Forbes

Data Journalism: How Big Data-Driven Analytics Improves Newsmaking

Machine-Learning Models Can Help Detect Early-Stage Cancer

PUBLISHED APRIL 22, 2022 IN ACADEMICS

TRACKING ATROCITIES USING BIG DATA

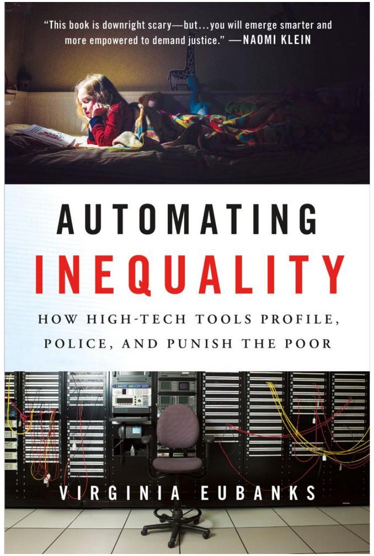
Data for Ukraine Project Shows Incidents As They Happen

The Guardian

Women less likely to be shown ads for high-paid jobs on Google, study shows

Samuel Gibbs

Wed 8 Jul 2015 06.29 EDT



Why? When it goes wrong, there are significant impacts on people's well-being.

Why me? The mechanisms for harm can be complex and difficult for non-data scientists to understand.

Equifax Agrees to $425 Million Breach Settlement

Company Settles Over 2017 Breach Affecting 147 Million People

Prajeet Nair (@prajeetspeaks) · February 9, 2022

Machine Bias

There's software used across the country to predict future criminals. And it's biased against blacks.

by Julia Angwin, Jeff Larson, Surya Mattu and Lauren Kirchner, ProPublica

May 23, 2016

PROPUBLICA

Common data ethics misconceptions

There is no one right answer. Tradeoffs are inevitable and reasonable people can disagree. Tradeoff calculations will depend on your specific use case and context.

Ethics need to be actively considered **throughout a project**, not just at the beginning

Good intention are not enough. To minimize the risk of harm, you have to intentionally consider possible consequences.

We'll talk about a **starting point** for incorporating ethics into **practical data science work.**

Deon is an open-source command line tool that allows you to easily add an ethics checklist to your data science projects.

Goal: We are all more intentional in our choices and more aware of the ethical implications of our work

deon.drivendata.org

What is deon?

Checklist

- [ ] **A.1 Informed consent**: If there are human subjects, have they given informed consent, where subjects affirmatively opt-in and have a clear understanding of the data uses to which they consent?
- [ ] **A.2 Collection bias**: Have we considered sources of bias that could be introduced during data collection and survey design and taken steps to mitigate those?
- [ ] **A.3 Limit PII exposure**: Have we considered ways to minimize exposure of personally identifiable information (PII) for example through anonymization or not collecting information that isn't relevant for analysis?

Checklist items are meant to provoke discussion

The goal of the checklist items are not to concretely recommend a specific action but rather are framed as prompts to discuss or consider.

Decisions on ethical courses of action are not up to data scientists alone.

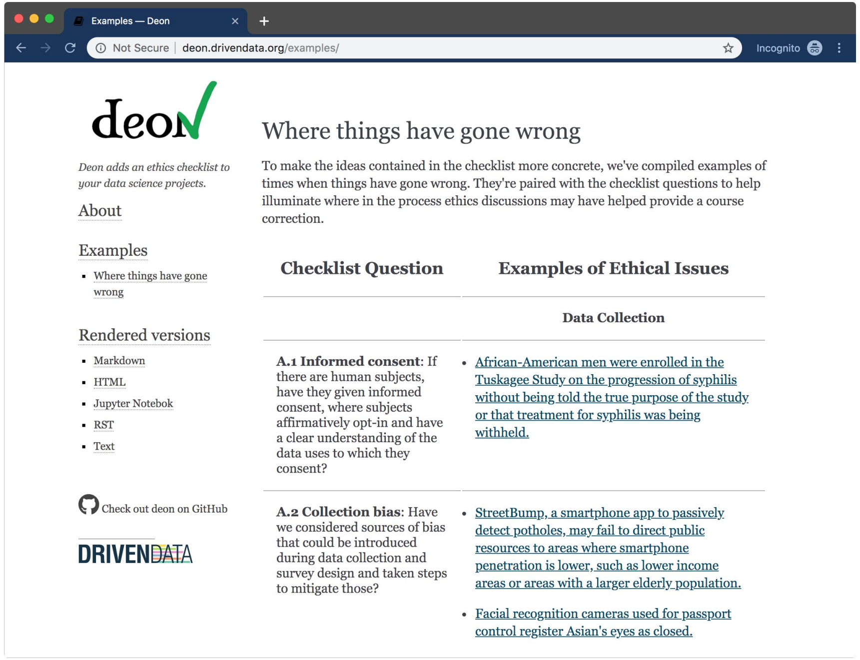
Checklist is designed to provoke conversations around issues where data scientists have particular responsibility and perspective.

Our goal is not to be arbitrators of what ethical concerns merit inclusion.

It is meant as a sensible starting point, and we believe teams will benefit from building custom checklists.

What is deon?

Examples



We believe in the power of examples to bring the principles of data ethics to bear on human experience.

The deon documentation includes a list of real-world examples connected with each item in the default checklist.

https://deon.drivendata.org/examples/

How to use deon

Example command line use

Write out the checklist to a new file, where you can add text and dig into each section:

```
$ deon --output ETHICS.md
```

Append the checklist to an existing Jupyter notebook (would also work with a markdown file):

```
$ deon --output existing-notebook-path.ipynb
```

Write out the checklist to a Jupyter notebook with multiple cells, and one checklist item per cell:

```
$ deon --output checklist.ipynb --multicell
```

Tips

You can also:

- Write to multiple **output file types** from the command line, including `.txt`, `.ipynb`, and more
- Create **custom checklists** for your own projects

To see all of the command line options:

```
$ deon --help
```

Five sections with distinct ethical risks

Data collection

Data storage

Analysis

Modeling

Deployment

We’ll now walk through the prompts for each stage and discuss examples.

Questions?

A. Data collection

Informed consent

If there are human subjects, have they given informed consent, where subjects affirmatively opt-in and have a clear understanding of the data uses to which they consent?

Collection bias

Have we considered sources of bias that could be introduced during data collection and survey design and taken steps to mitigate those?

continued...

A. Data collection

Limit PII exposure

Have we considered ways to minimize exposure of personally identifiable information (PII) for example through anonymization or not collecting information that isn't relevant for analysis?

Downstream bias mitigation

Have we considered ways to enable testing downstream results for biased outcomes (e.g., collecting data on protected group status like race or gender)?

A. Data collection

Informed consent: If there are human subjects, have they given informed consent?

Collection bias: Have we considered sources of bias that could be introduced during data collection?

Limit PII exposure: Have we considered ways to minimize exposure of personally identifiable information (PII)?

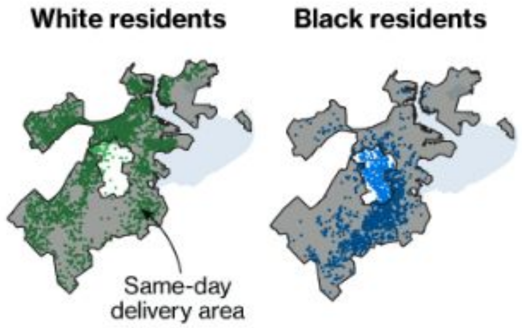
- **Downstream bias mitigation:** Have we considered ways to enable testing downstream results for biased outcomes?

Amazon Doesn't Consider the Race of Its Customers. Should It?

By David Ingold and Spencer Soper
April 21, 2016

Amazon's Prime Free Same-Day Delivery service initially **excluded predominantly black zip codes** in multiple major US cities. Amazon targeted areas with a high concentration of Prime members, and did not consider race data when drawing maps.

Three ZIP codes in the center of Boston, including the Roxbury neighborhood, are excluded from same-day coverage.



Source: Bloomberg 2016

B. Data storage

Data security

Do we have a plan to protect and secure data (e.g., encryption at rest and in transit, access controls on internal users and third parties, access logs, and up-to-date software)?

Right to be forgotten

Do we have a mechanism through which an individual can request their personal information be removed?

Data retention plan

Is there a schedule or plan to delete the data after it is no longer needed?

B. Data storage

- ## Data security

  Do we have a plan to protect and secure data (e.g., encryption at rest and in transit, access controls on internal users and third parties, access logs, and up-to-date software)?

  ## Right to be forgotten

  Do we have a mechanism through which an individual can request their personal information be removed?

  ## Data retention plan

  Is there a schedule or plan to delete the data after it is no longer needed?

Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach

Carole Cadwalladr *and* Emma Graham-Harrison

Sat 17 Mar 2018 18.03 EDT

"The data analytics firm that worked with Donald Trump's election team and the winning Brexit campaign harvested millions of Facebook profiles of US voters...and used them to build a powerful software program to predict and influence choices at the ballot box."

Any data you collect represents additional risk to users. There is a tradeoff between collecting data about protected classes to test your models for discrimination, and putting users at risk if your data security fails

Source: The Guardian 2018

C. Analysis

Missing perspectives

Have we sought to address blindspots in the analysis through engagement with relevant stakeholders (e.g., checking assumptions and discussing implications with affected communities and subject matter experts)?

Dataset bias

Have we examined the data for possible sources of bias and taken steps to mitigate or address these biases (e.g., stereotype perpetuation, confirmation bias, imbalanced classes, or omitted confounding variables)?

continued...

C. Analysis

Honest representation

Are our visualizations, summary statistics, and reports designed to honestly represent the underlying data?

Privacy in analysis

Have we ensured that data with PII are not used or displayed unless necessary for the analysis?

Auditability

Is the process of generating the analysis well documented and reproducible if we discover issues in the future?

C. Analysis

Missing perspectives: Have we sought to address blindspots in the analysis through engagement with relevant stakeholders

Dataset bias: Have we examined the data for possible sources of bias and taken steps to mitigate or address these biases

Honest representation: Are our visualizations, summary statistics, and reports designed to honestly represent the underlying data?

- **Privacy in analysis:** Have we ensured that data with PII are not used or displayed unless necessary for the analysis?

Auditability: Is the process of generating the analysis well documented and reproducible if we discover issues in the future?

Fitness tracking app Strava gives away location of secret US army bases

Alex Hern
@alexhern
Sun 28 Jan 2018 16.51 EST

A military base in Helmand Province, Afghanistan with route taken by joggers highlighted by Strava. Photograph: Strava Heatmap

Strava released a data visualization map of the activity tracked by its users, **without deanonymizing or removing sensitive location information** from users on military bases and spy outposts.

Source: The Guardian 2018

D. Modeling

Fairness across groups

Have we tested model results for fairness with respect to different affected groups (e.g., tested for disparate error rates)?

Metric selection

Have we considered the effects of optimizing for our defined metrics and considered additional metrics?

Explainability

Can we explain in understandable terms a decision the model made in cases where a justification is needed?

continued...

D. Modeling

Proxy discrimination

Have we ensured that the model does not rely on variables or proxies for variables that are unfairly discriminatory?

Communicate bias

Have we communicated the shortcomings, limitations, and biases of the model to relevant stakeholders in ways that can be generally understood?

D. Modeling

Fairness across groups: Have we tested model results for fairness with respect to different affected groups

Metric selection: Have we considered the effects of optimizing for our defined metrics

Explainability: Can we explain in understandable terms a decision the model made

- **Proxy discrimination:** Have we ensured that the model does not rely on variables or proxies for variables that are unfairly discriminatory?

Communicate bias: Have we communicated the shortcomings, limitations, and biases of the model to relevant stakeholders in ways that can be generally understood?

WILL KNIGHT BUSINESS NOV 19, 2019 9:15 AM

The Apple Card Didn't 'See' Gender—and That's the Problem

The way its algorithm determines credit lines makes the risk of bias more acute.

Apple's credit card offered far smaller lines of credit to women than men, even when other indicators of financial health were the same. No gender variable was explicitly included, but the algorithm learned to discriminate indirectly based on things that women were more likely to do

A model trained on data from a world with inequality will likely learn to be unequal

Source: Wired 2019

D. Modeling

Fairness across groups: Have we tested model results for fairness with respect to different affected groups

- **Metric selection:** Have we considered the effects of optimizing for our defined metrics

Explainability: Can we explain in understandable terms a decision the model made

Proxy discrimination: Have we ensured that the model does not rely on variables or proxies for variables that are unfairly discriminatory?

Communicate bias: Have we communicated the shortcomings, limitations, and biases of the model to relevant stakeholders in ways that can be generally understood?

The Facebook whistleblower says its algorithms are dangerous. Here's why.

By Karen Hao

October 5, 2021

Facebook's algorithms **prioritized engagement**, which tended to elevate controversial, extreme, or shocking posts that are often misinformation.

"Facebook ... knows—they have admitted in public—that engagement-based ranking is dangerous without integrity and security systems but then not rolled out those integrity and security systems in most of the languages in the world. It is pulling families apart. And in places like Ethiopia it is literally fanning ethnic violence" - whistleblower Francis Haugen

Source: MIT Technology Review 2021

E. Deployment

Monitoring and evaluation

How are we planning to monitor the model and its impacts after it is deployed (e.g., human review of high-stakes decisions, reviewing downstream impacts of errors or low-confidence decisions)?

Redress

Have we discussed with our organization a plan for response if users are harmed by the results?

Rollback

Is there a way to turn off or roll back the model in production if necessary?

Unintended use

Have we taken steps to identify and prevent unintended uses and abuse of the model and do we have a plan to monitor these once the model is deployed?

E. Deployment

- **Monitoring and evaluation**

  How are we planning to monitor the model and its impacts after it is deployed (e.g., human review of high-stakes decisions, reviewing downstream impacts of errors or low-confidence decisions)?

- **Redress**

  Have we discussed with our organization a plan for response if users are harmed by the results?

- **Rollback**

  Is there a way to turn off or roll back the model in production if necessary?

Unintended use

Have we taken steps to identify and prevent unintended uses and abuse of the model and do we have a plan to monitor these once the model is deployed?

WHAT HAPPENS WHEN AN ALGORITHM CUTS YOUR HEALTH CARE

By Colin Lecher | @colinlecher | Mar 21, 2018, 9:00am EDT

Software mistakes and the implementation of an algorithm resulted in drastic cuts to vital healthcare for people with cerebral palsy and other conditions. One example is a woman in Arkansas with cerebral palsy, whose care hours were cut from 56 hours per week to just 32.

No explanation of the cause was provided, and there was no simple appeal process to override the algorithm's decision or update the algorithm.

Source: The Verge 2018

Questions?

Break

10 min

Group activity

Qualitative case study

Scenario 1: Informative analysis

The London mayor’s office has hired your team to analyze public transit use in the city. They want to understand where there is unmet need for public transit and opportunity to reduce car traffic and lower greenhouse gas emissions.

Data

The mayor’s office has formed an agreement with Uber to get six months of ride data from the Uber app backend. For each ride, the data includes start location, end location, duration, price, information about the rider, information about the driver, and other metadata.

Task

Use this data to produce visualizations helping policymakers understand public transit needs in the city as well as trends across London neighborhoods.

Scenario 1: Informative analysis

The London mayor's office has hired your team to analyze public transit use in the city. They want to understand where there is unmet need for public transit and opportunity to reduce car traffic and lower greenhouse gas emissions.

Data

The mayor's office has formed an agreement with Uber to get six months of ride data from the Uber app backend. For each ride, the data includes start location, end location, duration, price, information about the rider, information about the driver, and other metadata.

Task

Use this data to produce visualizations helping policymakers understand public transit needs in the city as well as trends across London neighborhoods.

With your breakout group: *~10 min*

- Talk through each checklist item in the **data collection, storage,** and **analysis** sections of deon. Refer to the checklist online (deon.drivendata.org/#data-science-ethics-checklist)
- **Write 3-5 bullets per section** on a specific ethical concern you identified and how you could approach mitigating it (9-15 bullets total for each breakout group). Bullets don't have to cover every checklist item.
- Share these in the Teams channel for this session
- Choose one person from your group to share out your findings

Scenario 1: Informative analysis

The London mayor's office has hired your team to analyze public transit use in the city. They want to understand where there is unmet need for public transit and opportunity to reduce car traffic and lower greenhouse gas emissions.

Data

The mayor's office has formed an agreement with Uber to get six months of ride data from the Uber app backend. For each ride, the data includes start location, end location, duration, price, information about the rider, information about the driver, and other metadata.

Task

Use this data to produce visualizations helping policymakers understand public transit needs in the city as well as trends across London neighborhoods.

Share out!

A. Data Collection
A.1 Informed consent
A.2 Collection bias
A.3 Limit PII exposure
A.4 Downstream bias mitigation

B. Data Storage
B.1 Data security
B.2 Right to be forgotten
B.3 Data retention plan

C. Analysis
C.1 Missing perspectives
C.2 Dataset bias
C.3 Honest representation
C.4 Privacy in analysis
C.5 Auditability

Scenario 2: Decision making

A bank wants to lower the default rates on their loans so they hire you, a data scientist, to help them improve their creditworthiness assessments.

Data

Your dataset includes all the fields from previous loan applications:

Name	Occupation
Sex	Credit score
Education level	Savings account balance
Marital status	Desired loan amount
Income	Loan purpose

Task

Build a model to predict whether a given loan application should be approved.

Scenario 2: Decision making

A bank wants to lower the default rates on their loans so they hire you, a data scientist, to help them improve their creditworthiness assessments.

Data

Your dataset includes all the fields from previous loan applications:

Name	Occupation
Sex	Credit score
Education level	Savings account balance
Marital status	Desired loan amount
Income	Loan purpose

Task

Build a model to predict whether a given loan application should be approved.

With your breakout group: *~10 min*

- Talk through each checklist item in the **analysis, modeling,** and **deployment** sections of deon. Refer to the checklist online (deon.drivendata.org/#data-science-ethics-checklist)
- **Write 3-5 bullets per section** on a specific ethical concern you identified and how you could approach mitigating it (9-15 bullets total for each breakout group). Bullets don’t have to cover every checklist item.
- Share these in the Teams channel for this session
- Choose a **different person** from your group to share out your findings

Scenario 2: Decision making

A bank wants to lower the default rates on their loans so they hire you, a data scientist, to help them improve their creditworthiness assessments.

Data

Your dataset includes all the fields from previous loan applications:

Name	Occupation
Sex	Credit score
Education level	Savings account balance
Marital status	Desired loan amount
Income	Loan purpose

Task

Build a model to predict whether a given loan application should be approved.

Share out!

C. Analysis
C.1 Missing perspectives
C.2 Dataset bias
C.3 Honest representation
C.4 Privacy in analysis
C.5 Auditability

D. Modeling
D.1 Proxy discimination
D.2 Fairness across groups
D.3 Metric selection
D.4 Explainability
D.5 Communicate bias

E. Deployment
E.1 Monitoring and evaluation
E.2 Redress
E.3 Roll back
E.4 Unintended use

Break

~5 min

Hands-on coding exercise

Eviction data case study

Setup *(review)*

Review the set up instructions from the prerequisites worksheet

These are also listed on the homepage of the github repository: https://github.com/drivendataorg/msc-buckingham-data-ethics#installation-instructions

If you weren't able to complete these ahead of time and need help debugging, post in the Teams thread for this workshop.

1. Clone the workshop github repository

   ```
   git clone
   https://github.com/drivendataorg/msc-buckingham-data-ethics.git
   ```

2. Create a new environment. If you don't have conda installed, install Miniconda first: https://docs.conda.io/en/latest/miniconda.html

   ```
   conda create --name msc-buckingham-data-ethics -y python=3.8
   ```

3. Activate the new environment and install the python package requirements

   ```
   conda activate msc-buckingham-data-ethics
   cd msc-buckingham-data-ethics
   pip install -r requirements.txt
   ```

Activity instructions

1. Open your terminal
2. Navigate to the folder with your cloned version of the workshop repository, and make sure you have the most recent code:

   ```
   cd msc-buckingham-data-ethics
   git pull
   ```

3. Activate your environment: `conda activate msc-buckingham-data-ethics`
4. Launch a jupyter notebook: `jupyter notebook`
5. Open `notebooks/eviction-data-case-study.ipynb`. We'll walk through the notebook as a group, and break for a few independent coding exercises.

If you need help debugging during any of the exercises, post in the Teams thread for this workshop or send a direct message to one of the DrivenData team members. **We encourage you to collaborate and work together throughout the exercise!**

There is a more comprehensive version of the case study notebook in `notebooks/eviction-data-case-study-reference.ipynb`. You can refer to it if you are stumped during any of the coding exercises, but we strongly encourage solving problems on your own first!

Questions?

Thank you!

Learn more at https://deon.drivendata.org/

Workshop materials:
github.com/drivendataorg/msc-buckingham-data-ethics

Katie Wetstone, katie@drivendata.org
Isha Shah, isha@drivendata.org

DRIVENDATA